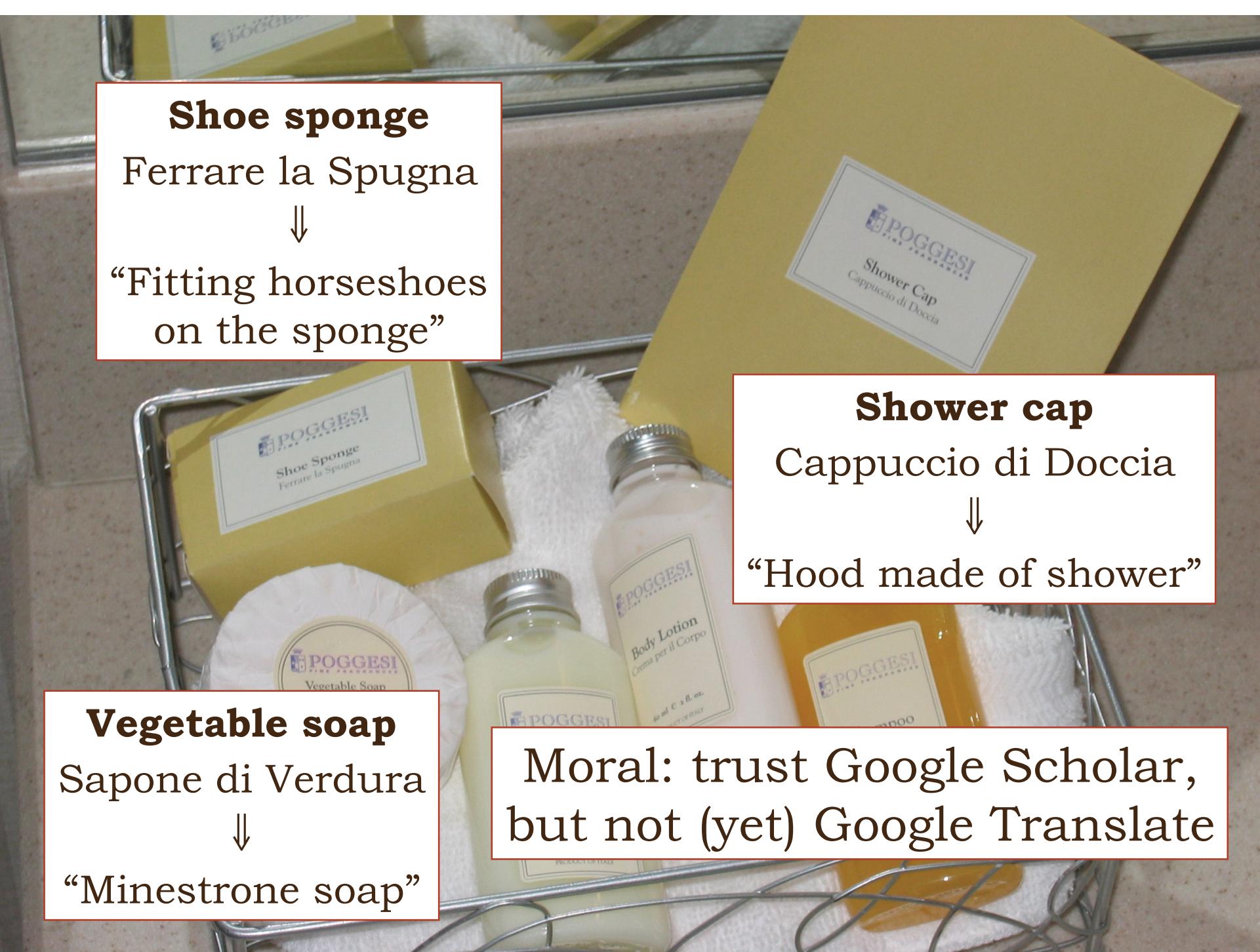




The Mock LISA Data Challenges: history, status, prospects

Michele Vallisneri (Jet Propulsion Laboratory)
for the MLDC Task Force:

Stas Babak, John Baker, Matt Benacquista,
Neil Cornish, Jeff Crowder, Curt Cutler, Shane Larson,
Tyson Littenberg, Edward Porter, M.V., Alberto Vecchio



Shoe sponge

Ferrare la Spugna



“Fitting horseshoes
on the sponge”

Vegetable soap

Sapone di Verdura



“Minestrone soap”

POGGESI
FINE FRAGRANCES

Shower Cap
Cappuccio di Doccia

Shower cap

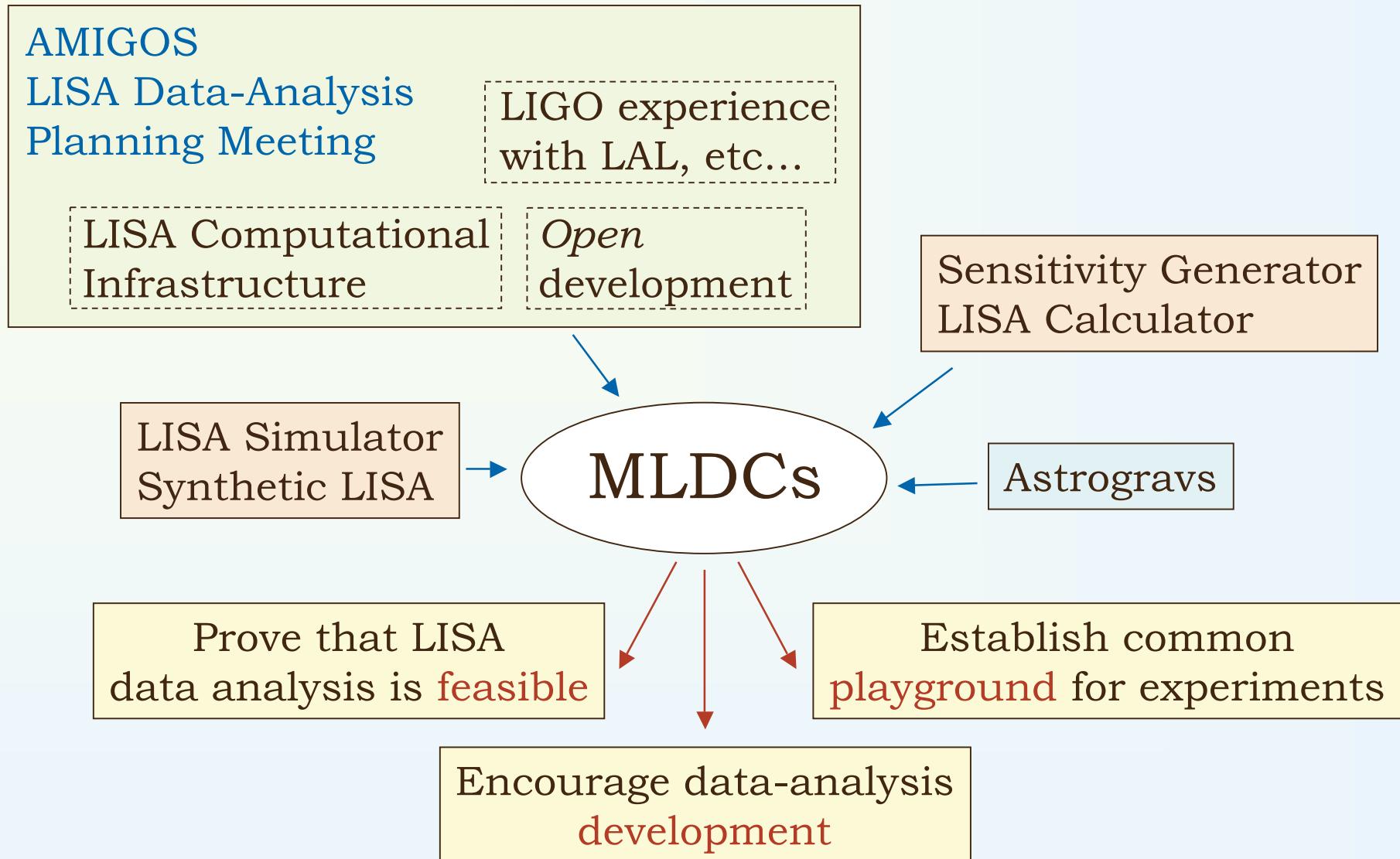
Cappuccio di Doccia



“Hood made of shower”

Moral: trust Google Scholar,
but not (yet) Google Translate

Genesis (fall 2005)



MLDCs: why?

- For LISA, data analysis is integral to the measurement concept
 - We must demonstrate that we can meet the LISA science requirements
 - We need to understand data analysis quantitatively to translate science requirements into design decisions
- Kickstart the development of a LISA data-analysis computational infrastructure
- Encourage, track, and compare progress in LISA data-analysis development in the open community

MLDCs: how?

- Coordinated, voluntary effort in GW community
- Periodically issue datasets with synthetic noise and GW signals from sources of undisclosed parameters; increasing difficulty
- Challenge participants return parameter estimates and descriptions of search methods

Mock LISA Data Challenge Task Force

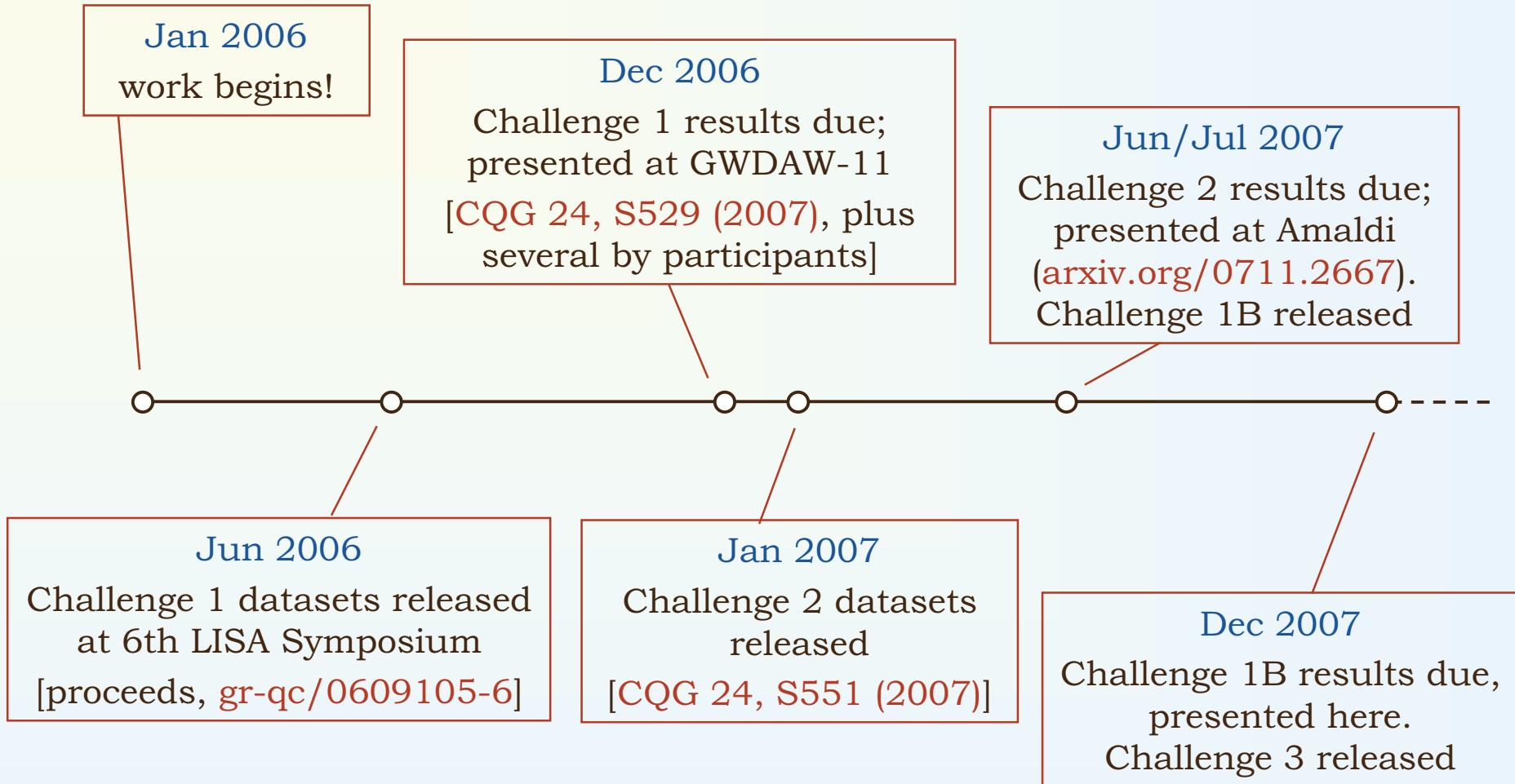
Members (past and present)

- Alberto Vecchio (co-chair)
- Michele Vallisneri (co-chair)
- Keith Arnaud
- Stas Babak
- John Baker
- Matt Benacquista
- Neil Cornish (WG-1B co-chair)
- Jeff Crowder
- Curt Cutler
- Sam Finn
- Steffen Grunewald
- Shane Larson
- Tyson Littenberg
- Eric Plagnol
- Ed Porter
- Sathyaprakash
- Jean-Yves Vinet

Charter

- Specify **pseudo-LISA** model
- Identify standard **source models**
- Specify **data format** (lisaXML)
- Plan challenge **progression**
- **Prepare and distribute** training and challenge datasets
- Develop **software infrastructure**
- **Compile** challenge submissions

MLDC timeline (so far)



	MLDC 1	MLDC 2	MLCD 1B	MLDC 3
Galactic binaries	<ul style="list-style-type: none"> • Verification ✓ • Unknown, isolated ✓ • Unknown, interfering ✓ 	<ul style="list-style-type: none"> • Galaxy of 3×10^7 ✓ 	<ul style="list-style-type: none"> • Verification ✓ • Unknown, isolated ✓ • Unknown, confused ✓ 	
MBH binaries	<ul style="list-style-type: none"> • Isolated ✓ 	<ul style="list-style-type: none"> • 4–6×, over Galaxy with EMRIs ✓ 	<ul style="list-style-type: none"> • Isolated ✓ 	
EMRIs		<ul style="list-style-type: none"> • Isolated ✓ • 4–6×, over Galaxy with SMBHs 	<ul style="list-style-type: none"> • Isolated ✓ 	
more...				

10 collaborations **13** collaborations **10** collaborations

Contestants...

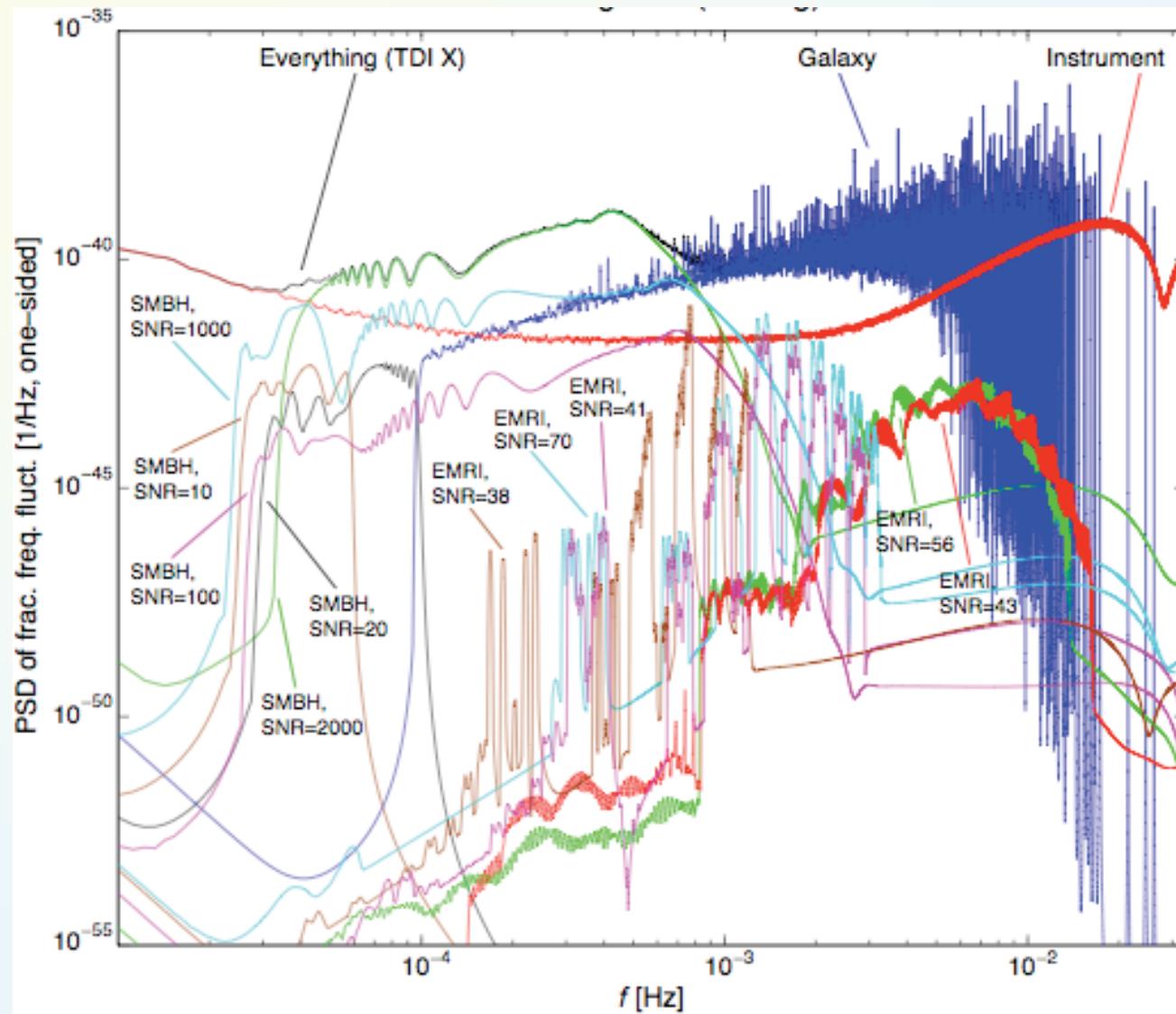
- NASA Ames
- U. of Auckland
- **Chinese Academy of Sci., Beijing**
- U. of Birmingham
- U. of Texas Brownsville
- Caltech/NASA JPL
- U. of Cambridge
- Cardiff U.
- Carleton College
- U. of Glasgow
- NASA Goddard
- Albert Einstein Institut Golm
- **Albert Einstein Institut Hannover**
- U. Illes Balears
- Indian Inst. of Tech., Kharagpur
- IMPAN Warszaw
- Montana State U.

- Nanjing U.
- CNRS Nice
- Northwestern U.
- CNRS APC Paris
- U. of Southampton
- **U. of Wroclaw**

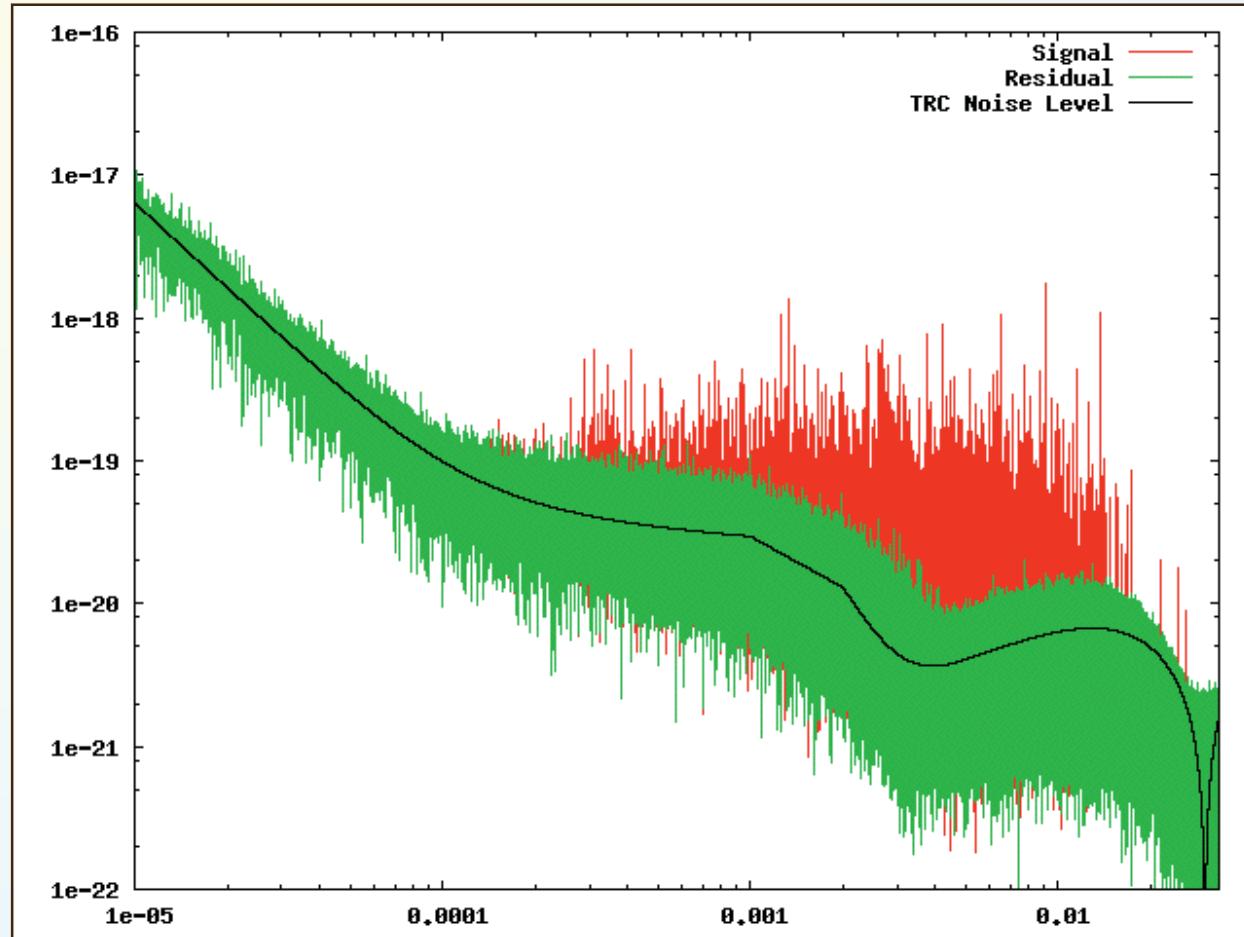
...and techniques

- Template-bank matched filtering
- Markov-Chain Monte Carlo matched filtering
- Genetic optimization
- Time-frequency track scans
- Tomographic reconstruction
- Hilbert transform
- F-statistic, hierarchical schemes
- ...

Challenge 2 highlights: the “Whole Enchilada”



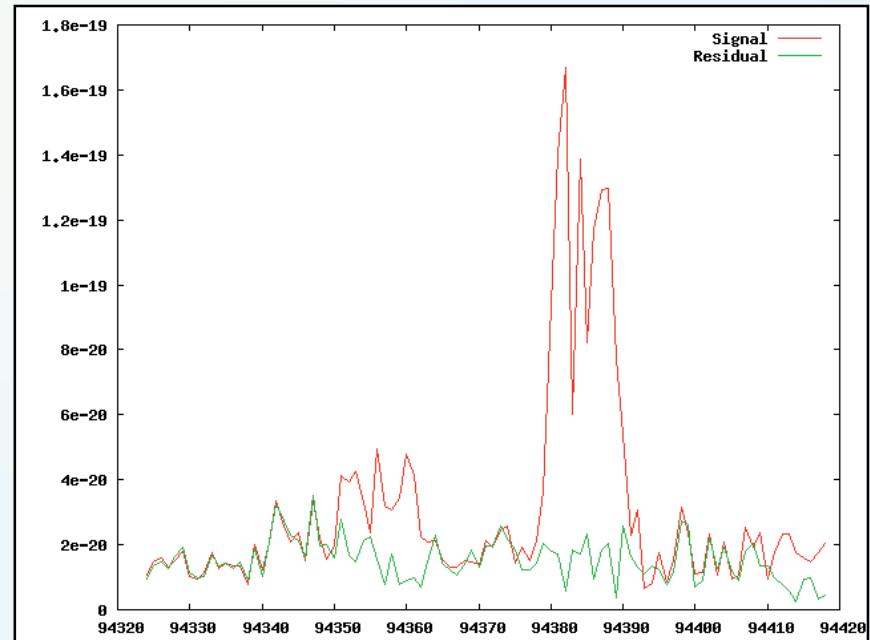
Challenge 2 highlights: Galaxy subtraction



- Using the MT/JPL catalog (thanks to Jeff Crowder) of 19324 sources, minus 1712 rejected by Bayesian Information Criterion

Challenge 1B results: Galactic binaries

- Several groups participated: AEI, Beijing/Nanjing, GSFC, IMPAN, Illes Balears/Birmingham
- Intrinsic parameters were recovered well, extrinsic parameters had some systematics (non-LW response)
- MLDC 1B.1.3 had no detectable sources (luckily, none were reported)
- GSFC and especially AEI did well on the interfering-binary dataset 1B.1.4
- AEI did OK on 1B.1.5 →
- Full results will be shown on MLDC website



Challenge 1B results: MBHs

MLDC 1B.2.1

- Caltech/JPL recovers 531.57/531.84 SNR, but it is on the wrong side of the sky (end-of-inspiral systematics)
- Cardiff recovers 511.77/531.84 SNR...

	$\Delta m_1/m_1$ ($\times 10^{-2}$)	$\Delta m_2/m_2$ ($\times 10^{-2}$)	$\Delta t_c/t_c$ ($\times 10^{-5}$)	$\Delta\beta$	$\Delta\lambda$	$\Delta D_L/D_L$ ($\times 10^{-1}$)	$\Delta\iota$ ($\times 10^{-1}$)	$\Delta\psi$	$\Delta\varphi_0$
Cardiff	12.1	10.01	3.601	1.374	0.549	5.89	6.87	4.835	-2.389
JPL	0.61	0.52	1.37	2.43	3.133	1.22	7.13	5.719	-2.846

MLDC 1B.2.2

- Caltech/JPL recovers 79.85/80.67 SNR, parameters OK.

	$\Delta m_1/m_1$ ($\times 10^{-1}$)	$\Delta m_2/m_2$ ($\times 10^{-1}$)	$\Delta t_c/t_c$ ($\times 10^{-5}$)	$\Delta\beta$ ($\times 10^{-3}$)	$\Delta\lambda$ ($\times 10^{-2}$)	$\Delta D_L/D_L$ ($\times 10^{-3}$)	$\Delta\iota$ ($\times 10^{-2}$)	$\Delta\psi$	$\Delta\varphi_0$
JPL	1.39	1.18	7.296	5.86	-1.462	4.803	-6.96	1.522	-4.725

Challenge 1B results: EMRIs

MLDC 1B.3.1

- Cornish recovers $123.36/123.65$ SNR
- Babak/Barak/Gair/Porter recovers $72.55/123.65$ SNR...
- Gair/Mandel/Wen do track search (no SNR)

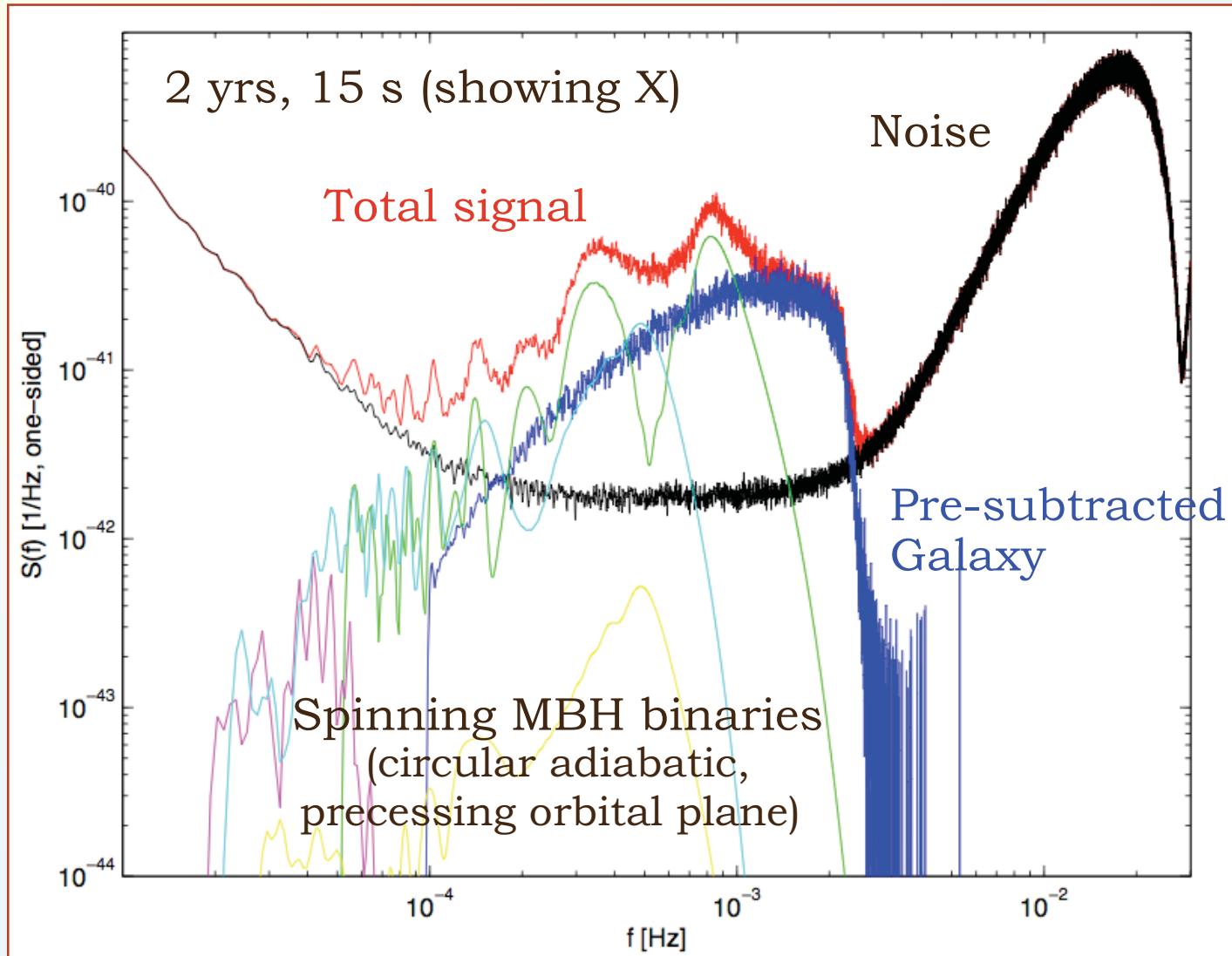
Entry	Index	$\frac{d\beta}{\Delta\beta}$	$\frac{d\lambda}{\Delta\lambda}$	$\frac{d\theta_K}{\Delta\theta_K}$	$\frac{d\phi_K}{\Delta\phi_K}$	$\frac{da}{\Delta a}$	$\frac{d\mu}{\Delta\mu}$	$\frac{dM}{\Delta M}$	$\frac{d\nu_0}{\nu_0}$	$\frac{d\Phi_0}{\Delta\Phi_0}$	$\frac{de_0}{0.15}$	$\frac{d\gamma_0}{\Delta\gamma_0}$	$\frac{d\alpha_0}{\Delta\alpha_0}$	$\frac{d\lambda_{SL}}{\Delta\lambda_{SL}}$	$\frac{dD}{D}$
BBGP-1B.3.1	1	-0.03	-0.0059	-0.14	0.053	0.31	-0.20	-0.84	0.026	0.11	0.37	-0.076	-0.43	-0.022	-1.62
EtfAG-1B.3.1	1	0.019	-0.0045	0.56	0.33	0.16	-0.11	-0.27	-9.3e-05	0.20	0.17	0.45	-0.018	0.078	-1.62
MT2-1B.3.1	1	0.0058	0.0027	0.00044	0.0051	-0.0022	0.0065	0.014	3.2e-06	-0.0048	-0.0085	0.48	0.014	-0.0020	-0.0076

	MLDC 1	MLDC 2	MLCD 1B	MLDC 3
Galactic binaries	<ul style="list-style-type: none"> • Verification ✓ • Unknown, isolated ✓ • Unknown, interfering ✓ 	<ul style="list-style-type: none"> • Galaxy of 3×10^6 ✓ 	<ul style="list-style-type: none"> • Verification ✓ • Unknown, isolated ✓ • Unknown, confused ✓ 	<ul style="list-style-type: none"> • Galaxy of 6×10^7 chirping
MBH binaries	<ul style="list-style-type: none"> • Isolated ✓ 	<ul style="list-style-type: none"> • 4–6×, over Galaxy with EMRIs ✓ 	<ul style="list-style-type: none"> • Isolated ✓ <p>pre-subtracted</p>	<ul style="list-style-type: none"> • Over Galaxy spinning, precessing
EMRIs		<ul style="list-style-type: none"> • Isolated ✓ • 4–6×, over Galaxy with SMBHs 	<ul style="list-style-type: none"> • Isolated ✓ 	<ul style="list-style-type: none"> • 4–6× together, weaker
more...			<p>raw observables, randomized noises</p>	<ul style="list-style-type: none"> • Cosmic string cusp bursts • Cosmological background
	10 collaborations	13 collaborations	10 collaborations	see you in 1 yr!

Challenge 3.1: the Galaxy

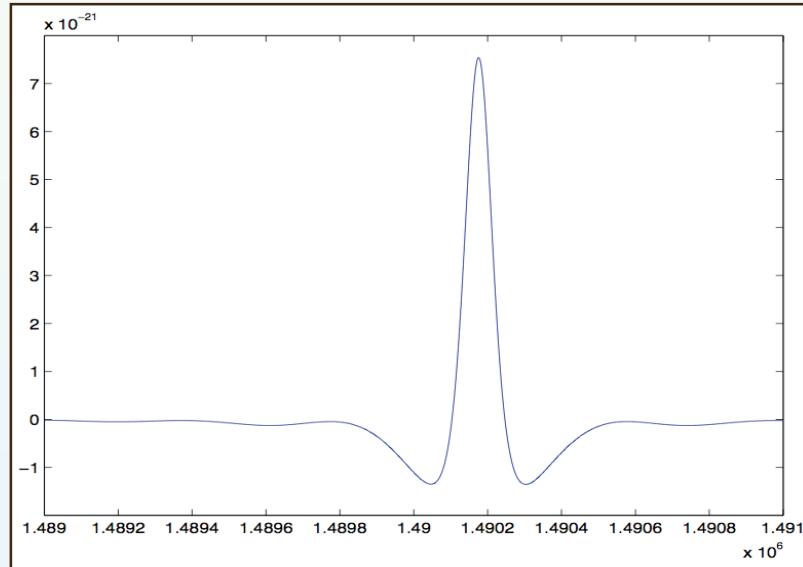
- Two years of data with 6×10^7 **Galactic binaries**
- Drawn from a randomized Nelemans population of 26 million **detached** and 34 million **interacting** binaries
- Orbital frequency increasing or decreasing, modeled as linear **chirps**
- 20 **verification binaries** of known position and frequency
- TDI X, Y, Z; secondary noise

Challenge 3.2: spinning BBHs



Challenge 3.3: EMRIs

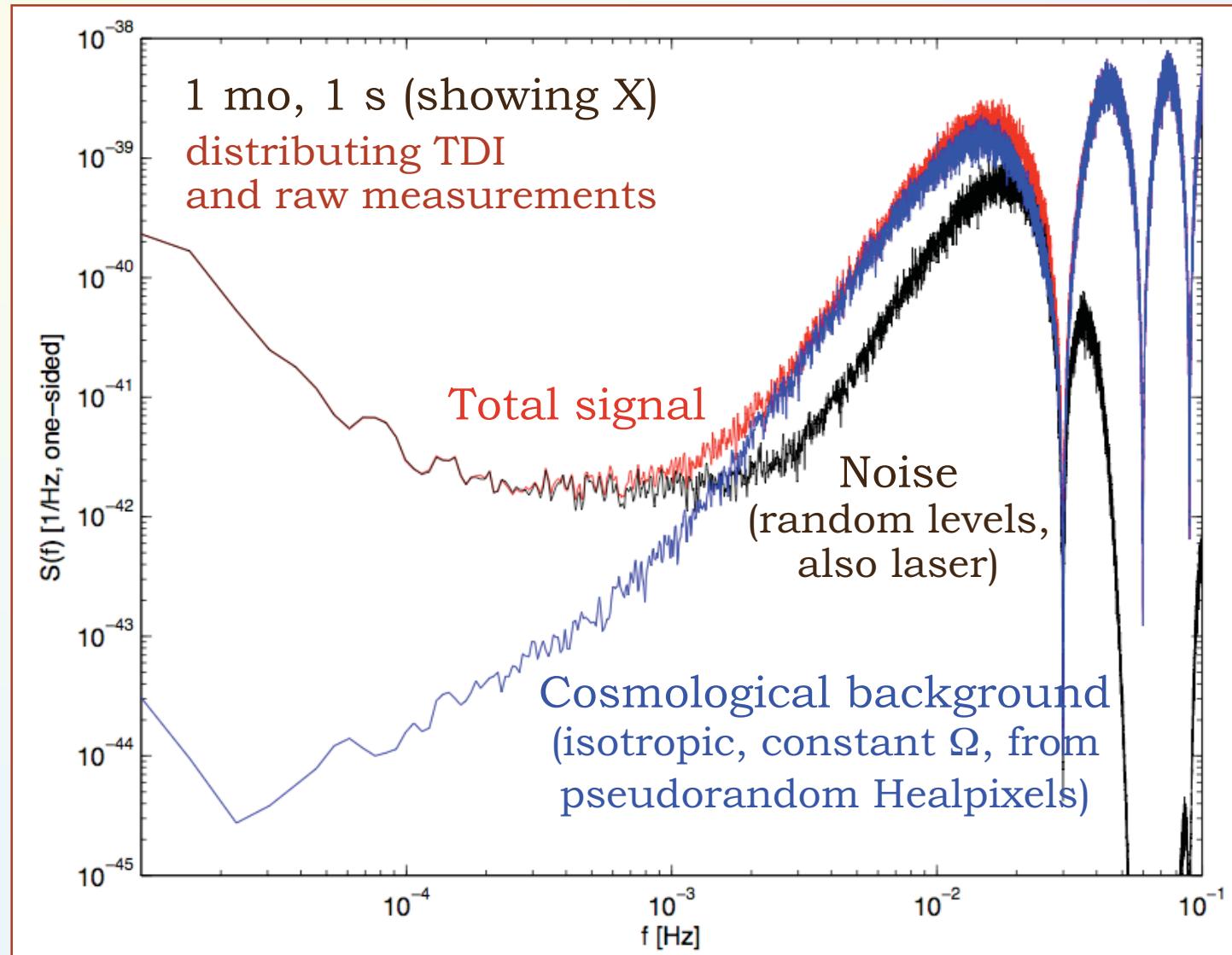
- Two years of data, 15 s, secondary noise only
- 5 EMRIs as in previous challenges, **superimposed**
- SNR 10–50



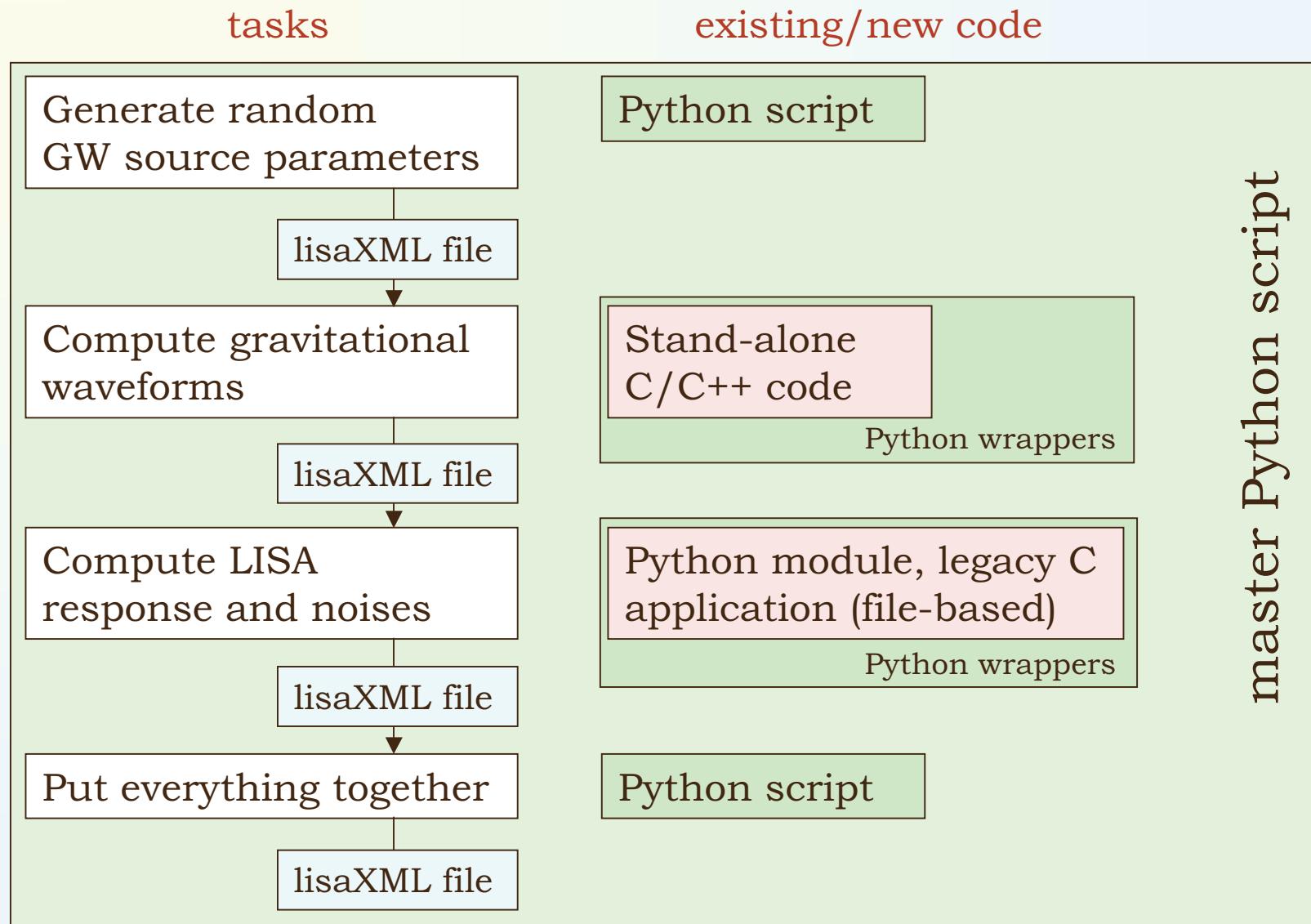
Ch. 3.4: cosmic strings

- 1 mo of data, 1 s, secondary noise + **low laser noise**
- TDI observables and **raw phase measurements**
- Poisson-distributed **cosmic-string cusp bursts** (LSC model)
- SNR 10–100

Challenge 3.5: cosmological background



The MLDC workflow



```

<?xml version="1.0" encoding="UTF-8"?>

<XSIL>
  <Param Name="LISA Data Source">
    <XSIL Type="List">
      [ ... ]
    </XSIL>
    <XSIL Type="Text">
      simulated data stream
    </XSIL>
    <XSIL Type="Text">
      binary data
    </XSIL>
  </Param>
</XSIL>
</XSIL>

```

Mock LISA Data Challenge XML File Format, v. 1.0

File Info

Authors	MLDC Task Force	
GenerationDate	2007-08-10T18:12:06CEST	ISO-8601

Full dataset for challenge1B.1.lc (synthlisa version), source seed = 733424, noise seed = 733424, LISAtools SVN revision 491 lisaXML 1.0 [M. Vallisneri, June 2006]

LISA data

Standard MLDC PseudoLISA (PseudoLISA)

TimeOffset	0	Second
InitialPosition	0	Radian
InitialRotation	0	Radian
Armlength	16.6782	Second

Source data

GB-1.1.1c (PlaneWave)

SourceType	GalacticBinary	
EclipticLatitude	-0.575706071762	Radian
EclipticLongitude	3.68595734709	Radian
Polarization	3.2062766975	Radian
Frequency	0.00974356389768	Hertz
InitialPhase	0.523693531091	Radian
Inclination	1.69786387662	Radian
Amplitude	1.98421310681e-23	1

TDI data

t,Xf,Yf,Zf (TDIObservable)

DataType	FractionalFrequency	
----------	---------------------	--

TimeSeries: t,Xf,Yf,Zf

TimeOffset	0.0	Second
Cadence	15.0	Second
Duration	31457280.0	Second
Array Stream: t,Xf,Yf,Zf	Filename	challenge1B.1.1c-training-frequency-0.bin
	Encoding	Binary, LittleEndian
	Type	double
	Unit	

lisaXML's natural Python interface

```
<?xml version="1.0"?>  
  
<XSIL>  
  <Param Name="Author">  
    Michele Vallisneri  
  </Param>  
  
  <XSIL Type="SourceData">  
    <XSIL Name="Galactic binary 1.1"  
          Type="PlaneWave">  
      <Param Name="SourceType">  
        GalacticBinary  
      </Param>  
      <Param Name="EclipticLatitude"  
            Unit="Radian">  
        0.9806443268  
      </Param>  
  
      [...more Params...]  
    </XSIL>  
  
    [...more PlaneWave sources...]  
  </XSIL>
```

```
load lisaXML file  
  
>>> fileobj = lisaXML('test.xml','r')  
>>> fileobj  
<lisaXML file 'test.xml'>  
  
>>> fileobj.Author  
  access metadata  
  'Michele Vallisneri'  
  
>>> fileobj.SourceData  
<XSIL SourceData (2 ch.)>  
  select XSIL  
  container  
  
>>> gb = fileobj.SourceData[0]  
>>> gb  
<XSIL PlaneWave 'Galactic binary 1.1'>  
  
>>> gb.Name  
  access  
  attributes  
  and Params  
  'Galactic binary 1.1.1a'  
>>> gb.EclipticLatitude  
  0.9806443268  
>>> gb.EclipticLatitude_Unit  
  'Radian'  
>>> gb.parameters  
  ['EclipticLatitude',  
   'EclipticLongitude', 'Polarization',  
   'Frequency', 'InitialPhase',  
   'Inclination', 'Amplitude']
```

In conclusion

- It's been a lot of work, but we're showing that LISA data analysis is possible, we're developing new techniques, we're publishing many papers (and see the poster session!)
- Cross-pollination with ground-based efforts is crucial
- The MLDC infrastructure (LISAtools) can be used to generate data for many other experiments outside the mainline challenges
- The LISA standard model (pseudo-LISA, source models) can be used to compare data-analysis results (see beginning investigations of LISA science performance)
- In the future: more realistic noise, sources; use MLDCs as testbed for prototypes of LISA core analysis tools

See for yourself

- MLDC official site:
astrogravs.nasa.gov/docs/mldc
- MLDC taskforce wiki:
www.tapir.caltech.edu/dokuwiki/listwg1b:home
- Mailing lists:
lisatools-mldc@gravity.psu.edu (formulation)
lisatools-challenge@gravity.psu.edu (participants)
- LISAtools software (including full MLDC pipeline):
lisatools.googlecode.com

When you're tired of analyzing real data,
come play with us!